

1 **Molecular methods enhance the detection of pyoderma-related *Streptococcus***
2 ***pyogenes* and *emm*-type distribution in children**

3

4 Jennifer N. Hall^{1,2,3}, Edwin P. Armitage^{4,5}, Elina Senghore⁴, Saffiatou Darboe⁴, Momodou
5 Barry⁴, Janko Camara⁴, Sulayman Bah⁴, Alexander J. Keeley^{1,2,4,5}, James S. McCarthy⁶, Pierre
6 Smeesters^{7,8}, Claire E. Turner^{2,3}, Thomas C. Darton^{1,2}, Michael Marks^{5,9,10}, Adrienn Angyal^{1,2},
7 Thushan I. de Silva^{1,2,4,5}

8

- 9 1. Division of Clinical Medicine, School of Medicine and Population Health, University of
10 Sheffield, Sheffield, United Kingdom
- 11 2. The Florey Institute, University of Sheffield, Sheffield, United Kingdom
- 12 3. School of Biosciences, University of Sheffield, Sheffield, United Kingdom
- 13 4. Vaccines and Immunity Theme, Medical Research Council Unit The Gambia at The
14 London School of Hygiene & Tropical Medicine, Banjul, PO Box 273, The Gambia
- 15 5. Department of Clinical Research, Faculty of Infectious and Tropical Diseases, London
16 School of Hygiene & Tropical Medicine, London, United Kingdom
- 17 6. Peter Doherty Institute of Infection and Immunity, University of Melbourne,
18 Melbourne, Australia
- 19 7. Department of Paediatrics, Brussels University Hospital, Academic Children Hospital
20 Queen Fabiola, Université libre de Bruxelles, 1020 Brussels, Belgium
- 21 8. Molecular Bacteriology Laboratory, Université Libre de Bruxelles, 1070 Brussels,
22 Belgium
- 23 9. Hospital for Tropical Diseases, University College London Hospital, London, United
24 Kingdom
- 25 10. Division of Infection & Immunity, University College London, London, United
26 Kingdom

27

28 **Corresponding author:** Jennifer Hall, Division of Clinical Medicine, School of Medicine and
29 Population Health, University of Sheffield, Sheffield, United Kingdom,
30 Jennifer.Hall@sheffield.ac.uk

31 **Manuscript word count:** 3077 words

32 **NOTE:** This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract word count: 200 words

33 **Abstract**

34

35 **Background:** *Streptococcus pyogenes*-related skin infections are increasingly implicated in the
36 development of rheumatic heart disease (RHD) in lower-resourced settings, where they are
37 often associated with scabies. The true prevalence of *S. pyogenes*-related pyoderma may be
38 underestimated by bacterial culture.

39

40 **Methods:** A multiplex qPCR for *S. pyogenes*, *Staphylococcus aureus* and *Sarcoptes scabiei* was
41 applied to 250 pyoderma swabs from a cross-sectional study of children <5 years in The
42 Gambia. Direct PCR-based *emm*-typing was used to supplement previous whole genome
43 sequencing (WGS) of cultured isolates.

44

45 **Results:** Pyoderma lesions with *S. pyogenes* increased from 51% (127/250) using culture to
46 80% (199/250) with qPCR. Compared to qPCR, the sensitivity of culture was 95.4% for *S.*
47 *pyogenes* (95% CI 77.2-99.9) in samples with *S. pyogenes* alone (22/250, 9%), but 59.9% (95%
48 CI 52.3-67.2) for samples with *S. aureus* co-infection (177/250, 71%). Direct PCR-based *emm*-
49 typing was successful in 50% (46/92) of cases, identifying 27 *emm*-types, including six not
50 identified by WGS (total 52 *emm*-types).

51

52 **Conclusions:** Bacterial culture significantly underestimates the burden of *S. pyogenes* in
53 pyoderma, particularly when co-infected with *S. aureus*. Molecular methods should be used
54 to enhance the detection of *S. pyogenes* in surveillance studies and clinical trials of
55 preventative measures in RHD-endemic settings.

56

57 **Key words:** *Streptococcus pyogenes*, The Gambia, skin infection, *emm*-typing, molecular
58 diagnostics, PCR, epidemiology, strain diversity

59

60

61

62

63

64

65 Introduction

66 Superficial bacterial skin infections (pyoderma) are common in childhood and ~162 million
67 children globally are affected at any one time [1]. A high prevalence is seen in low- and middle-
68 income countries, and in marginalised groups within high-income countries such as
69 indigenous populations [1–3]. *Streptococcus pyogenes* and *Staphylococcus aureus* are the
70 dominant causative pathogens, although microbiological data from pyoderma studies
71 conducted in these settings are scarce [1]. Previous studies have shown that polymerase chain
72 reaction (PCR) improves the detection of *S. pyogenes* in pharyngitis, but the use of molecular
73 assays for *S. pyogenes* in cases of pyoderma has been infrequently evaluated [4–6]. In addition
74 to a broad range of acute infections, post-infectious immune-mediated sequelae of *S.*
75 *pyogenes* include acute post-streptococcal glomerulonephritis and acute rheumatic fever
76 (ARF), leading to rheumatic heart disease (RHD) [7]. RHD following *S. pyogenes* pharyngitis is
77 well-described, but there is growing recognition of the role of *S. pyogenes* pyoderma in driving
78 RHD in settings with the highest burden of disease [8–10].

79

80 Pyoderma caused by both *S. pyogenes* and *S. aureus* is also frequently associated with
81 *Sarcoptes scabiei* infection (scabies) in RHD-endemic settings, which in turn is independently
82 associated with poverty and overcrowding [11–14]. The recent development of a PCR assay
83 to detect *S. scabiei* offers potential for a non-invasive, objective diagnosis of scabies infection
84 and integration alongside diagnostics for pyoderma [15].

85

86 Strain typing is an important component of epidemiological surveillance of *S. pyogenes*
87 infections, commonly carried out by sequencing the *emm* gene that encodes the M surface
88 protein [16]. The M protein has also been identified as a major *S. pyogenes* vaccine target,
89 therefore characterisation of *emm*-type distribution is essential to ensure adequate vaccine
90 coverage globally [17]. In high-income settings, the majority of *S. pyogenes* infections are
91 attributed to a small number of different *emm*-types, however, much greater genetic diversity
92 exists in resource-poor settings, with no clear *emm*-type dominance [18].

93

94 We previously conducted a cross-sectional study of 1441 children in The Gambia which
95 identified pyoderma in 250/1441 (17%) of children under five years old (95% CI 10–28).
96 Wound swab cultures yielded *S. aureus* in 81% and *S. pyogenes* in 51% of cases. Scabies

97 infection (diagnosed clinically) was seen in 16% of children (95% CI 12–20) and was
98 significantly associated with pyoderma [19]. Whole genome sequencing (WGS) of cultured
99 samples demonstrated diverse *emm*-types [20]. Here, we use a newly-established multiplex
100 quantitative PCR (qPCR) assay for *S. pyogenes*, *S. aureus* and *S. scabiei*, applied to wound swab
101 samples from this study, to explore whether the presence of these pathogens was
102 underestimated by clinical criteria and bacterial culture. Furthermore, we attempt to enhance
103 our data on strain diversity by using direct PCR-based *emm*-typing on *S. pyogenes* qPCR-
104 positive samples that were either culture negative, or where WGS of cultured isolates failed.

105

106 **Methods**

107 *Study design, sampling and bacterial culture*

108 Samples were taken during a cross-sectional, cluster-randomised, population-based study in
109 Sukuta, a peri-urban region in the Gambia, as previously described [19]. Ethical approval was
110 provided by the Gambia Government/MRC Joint Ethics Committee (SCC1587). Pyoderma was
111 defined as any skin lesion with evidence of pus or crusts. Children <5 years old within all
112 households in each cluster underwent skin examination by trained research nurses. Skin
113 lesions were clinically classified as scabies, infected scabies, pyoderma, or fungal skin
114 infection. Where pyoderma was diagnosed, superficial saline cleansing was done followed by
115 a single Nylon flocked swab (Copan) collected from the largest lesion into liquid Amie's
116 transport medium. Swabs were removed from transport media and inoculated the same day
117 on 5% sheep's blood agar and incubated overnight at 37°C. *S. aureus* and *S. pyogenes* were
118 identified through purity plates (where there was mixed infection), catalase, and
119 agglutination testing (Remel Staphaurex Plus or Streptex latex, Thermo Fisher Scientific).
120 Transport media were stored at -80°C until DNA extraction.

121

122 *DNA extraction and qPCR*

123 DNA was extracted from 500µL of transport media using the QIAmp DNA Mini Kit (Qiagen). A
124 pellet formed by centrifugation at 7000 xg for 5 minutes was resuspended in 180µL enzymatic
125 lysis buffer containing 40µL each of lysostaphin (1mg/ml) and lysozyme (100mg/ml). An
126 incubation at 37°C for 45 minutes was followed by addition of 25µL proteinase K and 200µL

127 AL buffer before a further incubation at 56°C for 60 minutes. DNA purification was then
128 carried out according to manufacturer instructions.

129 Bacterial loads were quantified using standard curves generated by eight 10-fold serial
130 dilutions of extracted DNA from *S. pyogenes* reference strain H293, *S. aureus* strain SH1000
131 and linearised plasmid DNA containing the *S. scabiei* SSR5 microsatellite sequence [15,21].
132 Dilutions ranged from 10,000,000 to 1 genome copy per PCR reaction.

133 Previously described *S. pyogenes* (*speB* gene), *S. aureus* (*nuc* gene), and *S. scabiei* (SSR5
134 microsatellite) PCR assays were integrated to establish a multiplex qPCR (Luna Universal
135 Probe qPCR Master Mix (New England Biolabs) and primers and probes as outlined in
136 **Supplementary table 1**) [15,21–23]. Nuclease-free water replicates were included as PCR
137 negative controls. Thermocycling conditions consisted of an initial ten minutes at 95°C,
138 followed by 40 amplification cycles of 94°C for 15 seconds and 58°C for 40 seconds. Limits of
139 detection (LOD) were determined using standard curves generated by eight 2.5-fold serial
140 dilutions from 1000 to 1.64 genome copies per PCR reaction, run in 11 replicates. The LOD
141 was defined as the lowest genome copy number that was amplified at a 95% detection rate
142 [24]. When tested in two replicates, the LOD was 31.1 copies for *S. pyogenes*, 3 copies for *S.*
143 *aureus*, and 4.5 copies for *S. scabiei*. Samples were run in duplicate, with the assay repeated
144 if high intra-replicate variation was seen (cycle threshold (Ct) standard deviation greater than
145 0.5). Amplification curves were reviewed to ensure consistency with true target amplification.
146 Samples were deemed positive if there was target amplification and a copy number above
147 the LOD.

148

149 *Direct PCR-based emm-typing*

150 The *emm* gene was amplified from swab DNA extracts using US Centers for Disease Control
151 and Prevention primers CDC1 (5'-TATTSGCTTAGAAAATTAA-3') and CDC2 (5'-
152 GCAAGTTCTTCAGCTTGTTT-3') [25]. PCR was performed with GoTaq polymerase (Promega) as
153 per manufacturer's instructions, with each reaction containing 200nM of the primers and 10µl
154 of DNA extract. Reaction conditions were at 94°C for 1 minute; followed by 30 cycles of 94°C
155 for 15s, 47°C for 30 seconds, 72°C for 85 seconds; and final extension at 72°C for 7 minutes.
156 PCR products were cleaned (Monarch PCR and DNA Cleanup Kit, New England Biolabs) and

157 eluted into 10µl, then used as the template for nested PCR using the same conditions and
158 primers CDC1 and CDC3 (5'-TTCTTCAAGCTCTTTGTT-3'). After gel visualisation, bands
159 corresponding to the *emm* gene (~1100bp) were purified (Monarch DNA Gel Extraction Kit,
160 New England Biolabs) and sent for Sanger sequencing (Genewiz, Azenta). If no band was seen
161 on gel visualisation, the first two PCR rounds were repeated from genomic DNA, followed by
162 a 3rd round nested PCR using high-fidelity polymerase Q5 (New England Biolabs) and primers
163 CDC1 and CDC2 (500nM each). Cycling conditions were 98°C for 30 seconds; followed by 34
164 cycles at 98°C for 10 seconds, 47°C for 30 seconds, 72°C for 30 seconds; and final elongation
165 at 72°C for 5 minutes.

166

167 *Emm*-types were assigned using the CDC *emm*-typing database tool [26]. An 180/180 base
168 match determined an *emm*-subtype; matches less than 180/180 but greater than 90/180 were
169 determined as *emm*-types.

170

171 *Data analysis*

172 Statistical analysis was performed using R Statistical Software (v4.2.2; R Core Team 2022).
173 Data were compared using a two-tailed Mann Whitney U test for continuous data or Fisher
174 exact test for categorical data. A p value of <0.05 was considered statistically significant, with
175 the Benjamini-Hochberg procedure applied for multiple hypothesis testing. Raw p values are
176 reported where they remain statistically significant after adjusting for multiple testing; others
177 are reported as not statistically significant. The Simpson Reciprocal Index of diversity was
178 calculated for samples with *emm*-types defined by WGS alone, and following addition of
179 direct-PCR typed samples [27,28].

180

181 **Results**

182 *Quantitative PCR shows greater diagnostic yield than bacteriological culture*

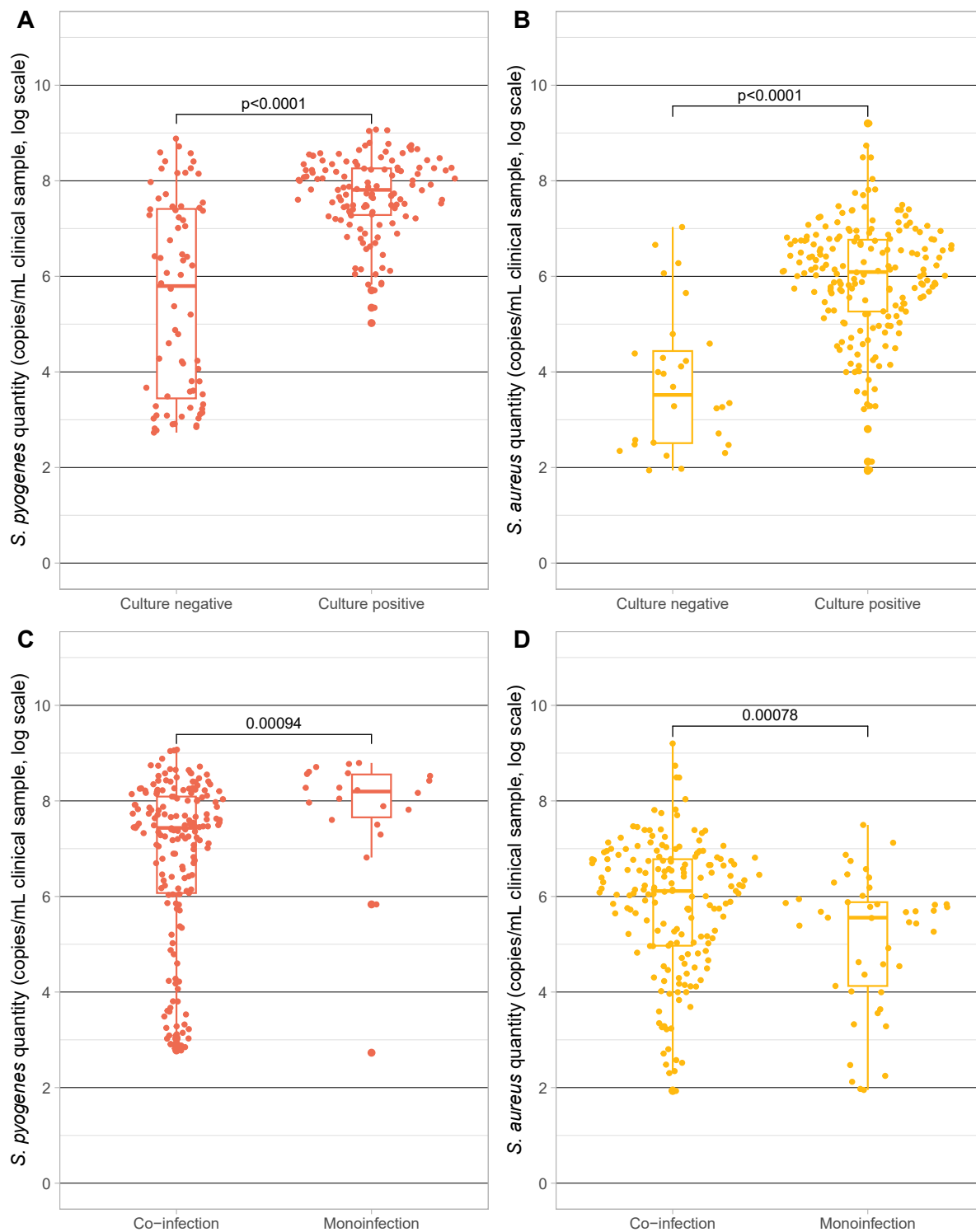
183 All transport media samples from pyoderma swabs underwent qPCR (n=250). One hundred
184 samples (40%) had either *S. pyogenes* or *S. aureus* detected by qPCR but not culture
185 (**Supplementary table 2**). The greatest additional diagnostic yield was in *S. pyogenes*,
186 identified in 72/250 samples (29%) by qPCR but not culture, resulting in 80% of pyoderma
187 cases being positive for *S. pyogenes*. All *S. pyogenes* culture-positive samples were also qPCR-
188 positive, but 8/250 samples (3%) that were *S. aureus* culture-positive were *S. aureus* qPCR-

189 negative. *S. aureus* and *S. pyogenes* co-infection was seen in 71% (177/250) of samples by
190 qPCR compared to 42% (104/250) by culture (**Supplementary figure 1**). Compared to qPCR,
191 the sensitivity of culture was 95.4% for *S. pyogenes* (95% CI 87-100) and 91.1% for *S. aureus*
192 (95% CI 83-99) in samples in which a single bacterial pathogen was identified, but 59.9% for
193 *S. pyogenes* (95% CI 53-67) and 86.4% for *S. aureus* (95% CI 81-91) from samples in which co-
194 infection was present.

195
196 For both *S. pyogenes* and *S. aureus*, qPCR bacterial load was significantly higher in samples
197 that were culture positive compared to those that were culture negative (both $p < 0.0001$)
198 (**Figure 1A and 1B**). *S. pyogenes* load in co-infected samples was significantly lower than in *S.*
199 *pyogenes* mono-infections ($p = 0.00094$), whereas *S. aureus* load was higher in co-infected
200 samples than in those with *S. aureus* alone ($p = 0.00078$, **Figure 1C and 1D**).

201
202 Forty three of 250 pyoderma swabs (17%) tested positive by qPCR for *S. scabiei*. Of the 25
203 swabs from lesions classified clinically as infected scabies, only four (16%) were positive by
204 qPCR (**Supplementary table 3**). In 45 cases where scabies was diagnosed clinically in another
205 body site, seven (16%) were qPCR-positive for *S. scabiei*, along with 18% (32/180) samples
206 from pyoderma cases where no clinical scabies diagnosis was made.

207
208
209



210

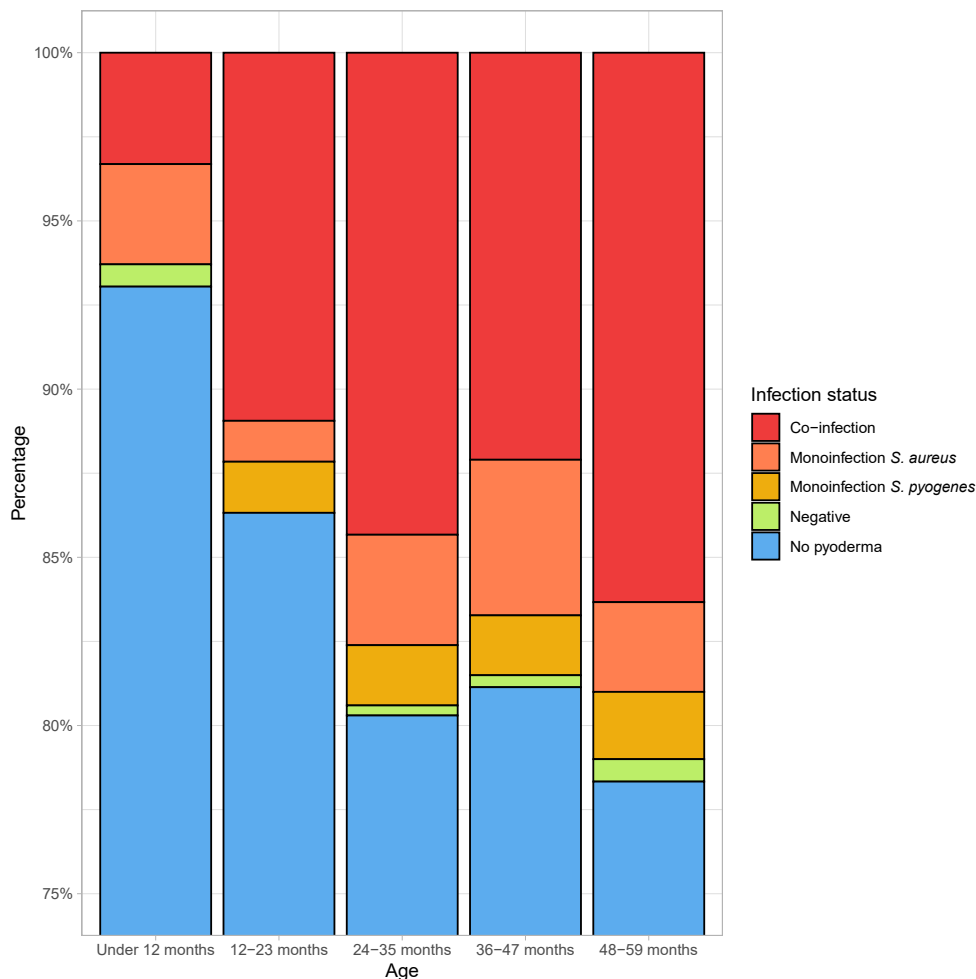
211 **Figure 1. Bacterial quantity in PCR-positive samples for (A) *S. pyogenes* and (B) *S. aureus* by**
212 ***bacteriological culture status; and for (C) *S. pyogenes* and (D) *S. aureus* in co-infected***
213 ***samples compared to samples with a single identified bacterial pathogen. Statistical***
214 ***differences determined using a two-tailed Mann Whitney U test.***

215

216

217 *Bacterial aetiology and load associations with age, body site and season*

218 The proportion of individuals with each pathogen increased with age (**Figure 2**). Within those
219 with pyoderma, age category was significantly associated with pattern of infection as defined
220 by qPCR result (Fisher exact test, $p = 0.039$), with no participants <1 year having infection with
221 *S. pyogenes* alone, and fewer participants having co-infection with both *S. pyogenes* and *S.*
222 *aureus* compared to older participants. Within all study participants, 3% (10/302) of those
223 under one year of age had pyoderma with *S. pyogenes*, rising to 13% (41/329) in those aged
224 12-24 months, and 18% (55/300) in those aged 48-59 months. Clinical sample bacterial load
225 of *S. pyogenes* and *S. aureus* did not significantly vary with participant age (**Supplementary**
226 **figure 2**).



227

228

229 **Figure 2. Participant infection status by age, as determined by qPCR result. Those with**
230 **infection status 'Negative' had a clinical diagnosis of pyoderma but neither *S. pyogenes* nor**
231 ***S. aureus* were detected by qPCR.**

232 Pyoderma was most commonly identified above the neck (50% of lesions), followed by leg or
233 foot (33%). Using qPCR, *S. pyogenes* was detected more commonly than *S. aureus* in the lower
234 limb (98% vs 85% of pyoderma lesions; **Supplementary figure 3A**). In contrast, in pyoderma
235 lesions above the neck, *S. aureus* was detected in 91% of lesions compared to *S. pyogenes* in
236 70%. There was a significant association between site of pyoderma and *S. pyogenes* detection
237 (Fisher exact test, $p < 0.0001$), but not *S. aureus* detection ($p = 0.28$). A greater *S. pyogenes* load
238 was seen in lower limb compared to above the neck lesions ($p < 0.0001$; **Supplementary figure**
239 **3B**), whereas no difference in *S. aureus* load was seen between these sites.

240

241 Our previous study reported a significant increase in the prevalence of pyoderma during the
242 rainy season at 23%, compared to 9% before the start of the rains. Co-infections, and mono-
243 infections with both *S. pyogenes* and *S. aureus* increased after the start of the rainy season
244 (**Figure 3A**), but the proportion of each infection type was not significantly different in cases
245 sampled before or after the start of the rainy season (**Figure 3B**). There was, however, a small
246 but statistically significant increase in both *S. pyogenes* and *S. aureus* bacterial loads in
247 pyoderma lesions sampled during the rainy season (**Figure 3C and 3D**).

248

249

250

251

252

253

254

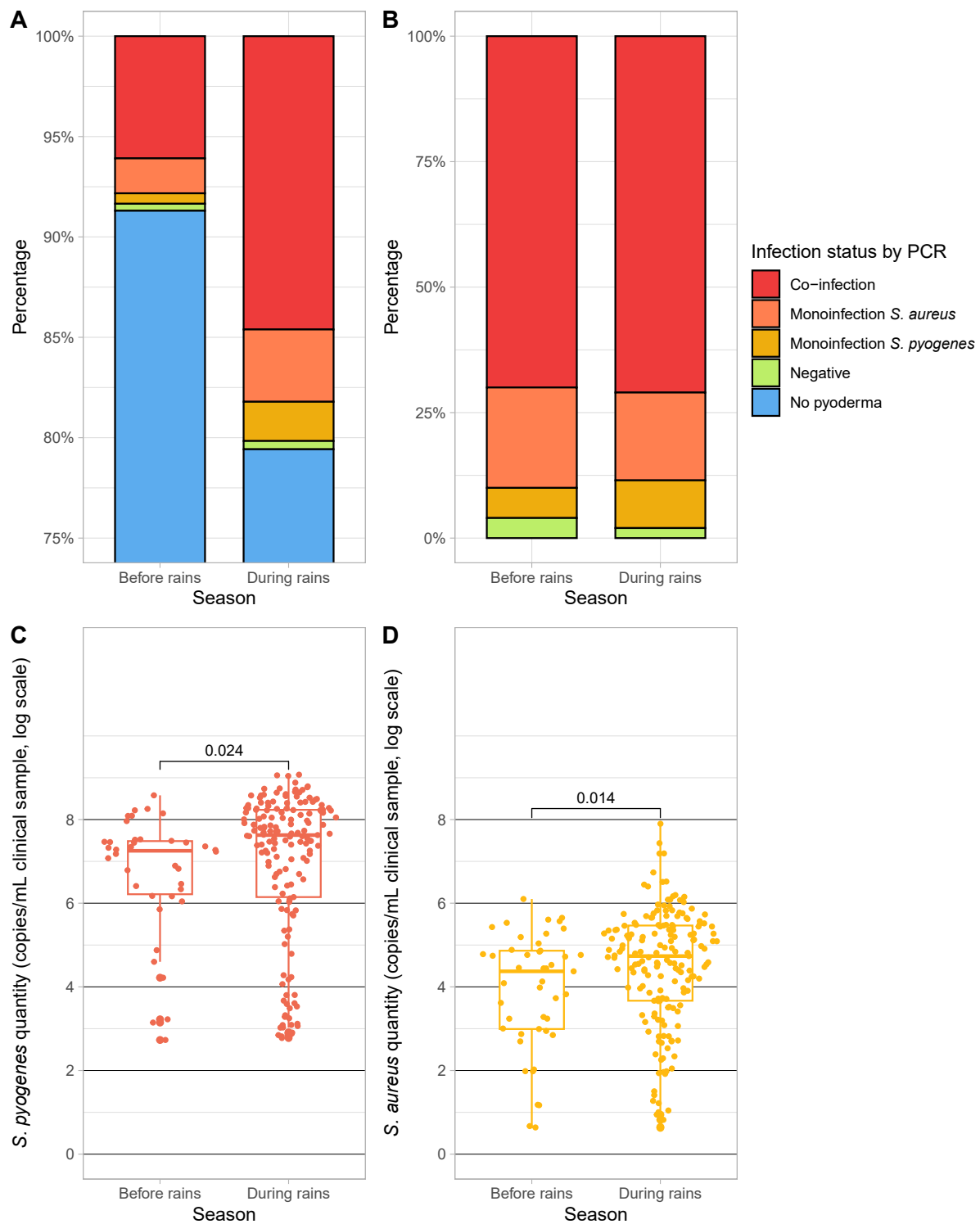
255

256

257

258

259



260

261

262

263

264

265

266

Figure 3. Bacterial aetiology and load of pyoderma cases according to season. Participant infection status by season, as determined by qPCR result, for (A) all study participants and (B) participants with pyoderma. Those with infection status 'Negative' had a clinical diagnosis of pyoderma but neither *S. pyogenes* nor *S. aureus* were detected by qPCR, whilst those with status 'No pyoderma' did not have swabs taken. Sample bacterial quantity for (C) *S. pyogenes* and (D) *S. aureus* in samples taken before the onset of the rainy season compared

267 to during the rainy season. Statistical differences determined using a two-tailed Mann
268 Whitney U test.

269

270 *Emm*-type distribution can be enhanced by direct PCR-based typing

271 Of the samples that were *S. pyogenes* qPCR positive, 127/199 were *S. pyogenes* culture
272 positive and previously underwent WGS. High quality sequence data were obtained for
273 107/127, from which 46 different *emm*-types were previously reported [20]. *Emm*-typing by
274 PCR was attempted on the remaining 92/199 samples, made up of 20/92 culture-positive
275 samples without high quality WGS data, and 72/92 which were *S. pyogenes* qPCR-positive but
276 culture-negative. *Emm*-types were successfully assigned by direct PCR-based typing for 46/92,
277 with 27 different *emm*-types identified, including six additional *emm*-types not identified by
278 WGS (*emm*66, 73, 82, 102, 111, 208) (**Supplementary table 4**). In *S. pyogenes* qPCR-positive
279 but culture-negative samples, the bacterial load was significantly higher in samples that were
280 successfully typed by direct PCR than untypeable samples ($p < 0.0001$) (**Supplementary figure**
281 **4**). The most common *emm*-type identified in the WGS dataset was *emm*80 (6/107), which
282 changed to *emm*4 (10/153) when the direct PCR-typed data were added (**Figure 4A**).

283

284 The Simpson Reciprocal Index of diversity was 49.3 (95% CI 39.0-66.9) for samples *emm*-typed
285 by WGS and 45.6 (95% CI 37.2-59.0) for all *emm*-typed samples. A high degree of *S. pyogenes*
286 *emm*-type diversity was seen within the nine sampled community geographical clusters
287 (**Figure 4B, Supplementary table 5**). Despite geographical proximity (within and between
288 households), the majority of *emm*-types were represented by a single isolate in a given
289 geographical cluster and limited enrichment for *emm*-types was seen.

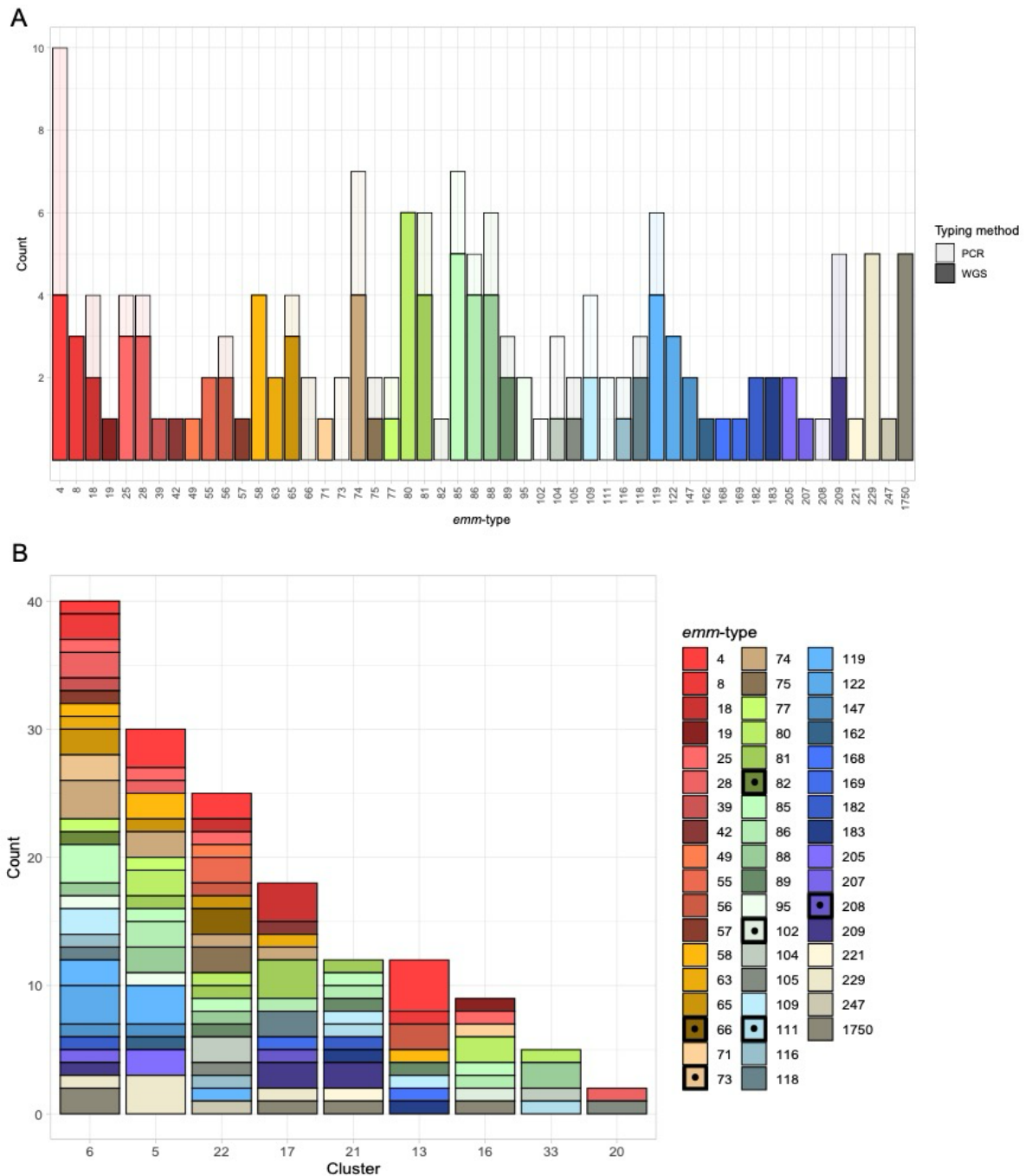
290

291

292

293

294



295

296

297 **Figure 4. Distribution of *S. pyogenes* emm-types.** Emm-types from 153 of 199 *S. pyogenes*
 298 qPCR positive samples were available. (A) Emm-types by typing method. Direct PCR-based
 299 emm-typing increased both the number of samples of 21 emm-types in the WGS-defined
 300 dataset, as well as detection of six emm-types not in the WGS dataset. (B) *S. pyogenes* emm-
 301 type by geographical cluster. Clusters ordered by detected number of pyoderma cases from
 302 high to low. Black boxes and dots in the emm-type legend denote emm-types identified by PCR
 303 but not WGS.

304 Discussion

305 In children <5 years old with pyoderma, we demonstrate a significant additional diagnostic
306 yield for *S. pyogenes* using a multiplex quantitative PCR assay in comparison to bacteriological
307 culture alone. Culture-based detection was most impacted when *S. aureus* co-infection was
308 present (sensitivity 59.9%), compared to high sensitivity (95.4%) in *S. pyogenes* mono-
309 infections. Samples with higher bacterial loads were more likely to be culture positive. Co-
310 infections with *S. pyogenes* and *S. aureus* were common (71% of pyoderma lesions), with
311 contrasting impact on bacterial load of the two organisms; lower *S. pyogenes* and higher *S.*
312 *aureus* load was seen in co-infections compared to mono-infections. The prevalence of both
313 organisms increased significantly with age and bacterial loads increased during the rainy
314 season. We also describe a novel method for *emm*-typing by PCR directly on patient samples
315 without intermediate culture. Direct PCR-based *emm*-typing in *S. pyogenes* qPCR-positive
316 cases where either culture or WGS had failed successfully generated 27 *emm*-types, further
317 enhancing the diverse set of *emm*-types previously generated by WGS.

318

319 In studies of pharyngitis, PCR has been demonstrated to have greater yield in the detection of
320 *S. pyogenes* when compared to bacteriological culture [4,5]. Our study shows that culture also
321 underestimates *S. pyogenes* infection in pyoderma lesions, with qPCR identifying *S. pyogenes*
322 in 80% of samples compared to 51% of samples by culture. High rates of co-infection with *S.*
323 *pyogenes* and *S.aureus* have been described previously, with culture-based studies in First
324 Nations populations in Australia or Canada reporting rates from 29% to 58% [10,29]. We show
325 that most *S. pyogenes* missed by culture are in co-infected samples. Although overgrowth of
326 *S. aureus* during culture and missed *S. pyogenes* colonies may explain this finding, we also find
327 that bacterial loads of *S. pyogenes* prior to culture are lower when *S. aureus* is present.
328 Whether this reflects bacterial inhibition and competition warrants further investigation.

329

330 Our original study reported a rise in pyoderma prevalence with age, seen in 7% of those under
331 one year old compared to 21% of those aged three to four years. With more complete
332 detection using qPCR, we demonstrate that *S. aureus* infections are more common at a
333 younger age, with mono- and co-infections with *S. pyogenes* increasing in later childhood.
334 More lower limb lesions were infected with *S. pyogenes* compared to other body sites. This
335 may reflect carriage distribution and transmission patterns, with *S. pyogenes* lower limb

336 pyoderma resulting from behavioural factors and child-to-child transmission. Similarly, *S.*
337 *aureus* was more common in lesions above the neck, which again could reflect proximity to
338 where *S. aureus* carriage burden may be highest in the nose.

339

340 There were considerable differences in our *S. scabiei* qPCR results in comparison to clinical
341 diagnosis, which may be explained by several factors. DNA extraction conditions we used
342 were chosen to enhance extraction from Gram-positive bacteria rather than arthropods
343 [15,30]. Swabs were taken of the pus and crusts associated with pyoderma lesions rather than
344 from skin lesions clinically typical of scabies infection. Most importantly, we compared qPCR
345 result to clinical diagnosis, rather than to a more accurate diagnostic method such as
346 microscopy. Previous studies have highlighted significant variation in clinical diagnoses in
347 both expert and non-expert examiners, and it is possible there are discrepancies between
348 clinical diagnoses and true number of scabies lesions [31,32].

349

350 *S. pyogenes* strain diversity was high in our study, with WGS and PCR-based *emm*-typing
351 identifying 52 different *emm*-types across 153 of our isolates. We have previously
352 demonstrated much higher genetic diversity in skin infection isolates from the Gambia
353 compared to the UK [20]. Limited *emm*-type enrichment was seen within geographical
354 clusters, further emphasising the diversity of circulating *emm*-types that are transmitted
355 between children in this setting. This pattern is reflected globally with a higher diversity of
356 *emm*-types seen in low-income settings and may have implications for pathogenesis of RHD,
357 with repeated exposures to different *emm*-types central to auto-immune priming [18,33,34].

358

359 We have previously outlined some of the limitations related to study design and sampling in
360 of our original study [19]. The likelihood of bacterial detection and accurate loads are reliant
361 on the quality of the swabbing. Samples had been archived at -80°C for several years prior to
362 undergoing qPCR, with the potential for DNA degradation during that time. A small number
363 of samples were positive for *S. aureus* by culture but not on qPCR, suggesting laboratory
364 contamination of culture plates, misidentification of cultured isolates, or qPCR failure in these
365 samples.

366

367 Our findings have implications for the design of interventions against pyoderma, both to
368 reduce the burden of skin disease but also as a constituent of the strategy to address RHD.
369 With the emerging recognition of the role of *S. pyogenes* pyoderma in the development of
370 ARF and RHD, understanding whether low-burden detection of *S. pyogenes* in skin lesions
371 may contribute to the immune priming implicated in RHD pathogenesis will be essential [8–
372 10,34–36]. With restored global interest in the development of a vaccine against *S. pyogenes*
373 and RHD [37], it is critical to have a robust understanding of the epidemiology and strain
374 diversity of *S. pyogenes* infection. Molecular methods should be central in enhanced
375 surveillance for *S. pyogenes* in high-burden settings to aid the design and assessment of future
376 interventions against RHD.

377

378

379 **References**

- 380 1. Bowen AC, Mahé A, Hay RJ, et al. The Global Epidemiology of Impetigo: A Systematic
381 Review of the Population Prevalence of Impetigo and Pyoderma. Reid SD, editor. PLOS
382 ONE. **2015**; 10(8):e0136789.
- 383 2. Romani L, Steer AC, Whitfeld MJ, Kaldor JM. Prevalence of scabies and impetigo
384 worldwide: a systematic review. Lancet Infect Dis. **2015**; 15(8):960–967.
- 385 3. Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A
386 streptococcal diseases. Lancet Infect Dis. **2005**; 5(11):685–694.
- 387 4. Henson AM, Carter D, Todd K, Shulman ST, Zheng X. Detection of Streptococcus
388 pyogenes by Use of Illumigene Group A Streptococcus Assay. J Clin Microbiol. American
389 Society for Microbiology; **2020**; 51(12):4207–4209.
- 390 5. Uhl JR, Adamson SC, Vetter EA, et al. Comparison of LightCycler PCR, Rapid Antigen
391 Immunoassay, and Culture for Detection of Group A Streptococci from Throat Swabs. J
392 Clin Microbiol. American Society for Microbiology; **2003**; 41(1):242–249.

- 393 6. Close RM, Sutcliffe CG, Galdun P, et al. Point-of-care molecular diagnostics for the
394 detection of group A Streptococcus in non-invasive skin and soft tissue infections: a
395 validation study. *Diagn Microbiol Infect Dis.* **2022**; 103(4):115729.
- 396 7. Walker MJ, Barnett TC, McArthur JD, et al. Disease Manifestations and Pathogenic
397 Mechanisms of Group A Streptococcus. *Clin Microbiol Rev.* **2014**; 27(2):264–301.
- 398 8. Oliver J, Bennett J, Thomas S, et al. Preceding group A streptococcus skin and throat
399 infections are individually associated with acute rheumatic fever: evidence from New
400 Zealand. *BMJ Glob Health. BMJ Specialist Journals;* **2021**; 6(12):e007038.
- 401 9. O’Sullivan L, Moreland NJ, Webb RH, Upton A, Wilson NJ. Acute Rheumatic Fever After
402 Group A Streptococcus Pyoderma and Group G Streptococcus Pharyngitis. *Pediatr*
403 *Infect Dis J.* **2017**; 36(7):692.
- 404 10. McDonald MI, Towers RJ, Andrews RM, Bengner N, Currie BJ, Carapetis JR. Low Rates of
405 Streptococcal Pharyngitis and High Rates of Pyoderma in Australian Aboriginal
406 Communities Where Acute Rheumatic Fever Is Hyperendemic. *Clin Infect Dis.* **2006**;
407 43(6):683–689.
- 408 11. World Health Organization. Ending the neglect to attain the Sustainable Development
409 Goals: a strategic framework for integrated control and management of skin-related
410 neglected tropical diseases. Geneva, Switzerland; 2022.
- 411 12. Vos T, Allen C, Arora M, et al. Global, regional, and national incidence, prevalence, and
412 years lived with disability for 310 diseases and injuries, 1990–2015: a systematic
413 analysis for the Global Burden of Disease Study 2015. *The Lancet.* **2016**;
414 388(10053):1545–1602.
- 415 13. Karimkhani C, Colombara DV, Drucker AM, et al. The global burden of scabies: a cross-
416 sectional analysis from the Global Burden of Disease Study 2015. *Lancet Infect Dis.*
417 **2017**; 17(12):1247–1254.
- 418 14. Hay RJ, Steer AC, Engelman D, Walton S. Scabies in the developing world—its prevalence,
419 complications, and management. *Clin Microbiol Infect.* **2012**; 18(4):313–323.

- 420 15. Chng L, Holt DC, Field M, et al. Molecular diagnosis of scabies using a novel probe-based
421 polymerase chain reaction assay targeting high-copy number repetitive sequences in
422 the *Sarcoptes scabiei* genome. *PLoS Negl Trop Dis*. **2021**; 15(2):e0009149.
- 423 16. Enright MC, Spratt BG, Kalia A, Cross JH, Bessen DE. Multilocus sequence typing of
424 *Streptococcus pyogenes* and the relationships between emm type and clone. *Infect*
425 *Immun*. **2001**; 69(4):2416–2427.
- 426 17. Vekemans J, Gouvea-Reis F, Kim JH, et al. The Path to Group A *Streptococcus* Vaccines:
427 World Health Organization Research and Development Technology Roadmap and
428 Preferred Product Characteristics. *Clin Infect Dis*. **2019**; 69(5):877–883.
- 429 18. Smeesters PR, Crombrughe G de, Tsoi SK, et al. Global *Streptococcus pyogenes* strain
430 diversity, disease associations, and implications for vaccine development: a systematic
431 review. *Lancet Microbe* [Internet]. **2023** [cited 2023 Dec 8]; . Available from:
432 <https://www.sciencedirect.com/science/article/pii/S266652472300318X>
- 433 19. Armitage EP, Senghore E, Darboe S, et al. High burden and seasonal variation of
434 paediatric scabies and pyoderma prevalence in The Gambia: A cross-sectional study.
435 *PLoS Negl Trop Dis*. Public Library of Science; **2019**; 13(10):e0007801.
- 436 20. Bah SY, Keeley AJ, Armitage EP, et al. Genomic Characterization of Skin and Soft Tissue
437 *Streptococcus pyogenes* Isolates from a Low-Income and a High-Income Setting. Torres
438 AG, editor. *mSphere*. **2023**; 8(1):e00469-22.
- 439 21. Keeley AJ, Groves D, Armitage EP, et al. *Streptococcus pyogenes* Colonization in Children
440 Aged 24–59 Months in the Gambia: Impact of Live Attenuated Influenza Vaccine and
441 Associated Serological Responses. *J Infect Dis*. **2023**; 228(7):957–965.
- 442 22. Dunne EM, Marshall JL, Baker CA, et al. Detection of group a streptococcal pharyngitis by
443 quantitative PCR. *BMC Infect Dis*. **2013**; 13(1):312.
- 444 23. Chesneau O, Allignet J, Solh N el. Thermonuclease gene as a target nucleotide sequence
445 for specific recognition of *Staphylococcus aureus*. *Mol Cell Probes*. **1993**; 7(4):301–310.

- 446 24. Klymus KE, Merkes CM, Allison MJ, et al. Reporting the limits of detection and
447 quantification for environmental DNA assays. *Environ DNA*. **2020**; 2(3):271–282.
- 448 25. Streptococcus Lab: Protocol for emm Typing | CDC [Internet]. 2023 [cited 2023 Nov 23].
449 Available from: <https://www.cdc.gov/streplab/groupa-strep/emm-typing-protocol.html>
- 450 26. Streptococcus Lab | StrepLab | Blast-emm GAS Subtyping Request Form | CDC
451 [Internet]. 2018 [cited 2024 Jan 10]. Available from:
452 <https://www2.cdc.gov/vaccines/biotech/>
- 453 27. Peet RK. The Measurement of Species Diversity. *Annu Rev Ecol Syst*. Annual Reviews;
454 **1974**; 5:285–307.
- 455 28. Grundmann H, Hori S, Tanner G. Determining Confidence Intervals When Measuring
456 Genetic Diversity and the Discriminatory Abilities of Typing Methods for
457 Microorganisms. *J Clin Microbiol*. **2001**; 39(11):4190–4192.
- 458 29. Bowen AC, Tong SY, Chatfield MD, Carapetis JR. The microbiology of impetigo in
459 Indigenous children: associations between *Streptococcus pyogenes*, *Staphylococcus*
460 *aureus*, scabies, and nasal carriage. *BMC Infect Dis*. **2014**; 14(1):727.
- 461 30. QIAamp DNA Mini Handbook.pdf.
- 462 31. Steer AC, Tikoduadua LV, Manalac EM, Colquhoun S, Carapetis JR, Maclennan C.
463 Validation of an Integrated Management of Childhood Illness algorithm for managing
464 common skin conditions in Fiji. *Bull World Health Organ*. **2009**; 87(3):173–179.
- 465 32. Osti MH, Sokana O, Gorae C, Whitfeld MJ, Steer AC, Engelman D. The diagnosis of scabies
466 by non-expert examiners: A study of diagnostic accuracy. *PLoS Negl Trop Dis*. Public
467 Library of Science; **2019**; 13(8):e0007635.
- 468 33. Steer AC, Law I, Matatolu L, Beall BW, Carapetis JR. Global emm type distribution of
469 group A streptococci: systematic review and implications for vaccine development.
470 *Lancet Infect Dis*. **2009**; 9(10):611–616.

- 471 34. Lorenz N, Ho TKC, McGregor R, et al. Serological Profiling of Group A Streptococcus
472 Infections in Acute Rheumatic Fever. *Clin Infect Dis.* **2021**; 73(12):2322–2325.
- 473 35. Whitcombe AL, McGregor R, Bennett J, et al. Increased Breadth of Group A
474 Streptococcus Antibody Responses in Children With Acute Rheumatic Fever Compared
475 to Precursor Pharyngitis and Skin Infections. *J Infect Dis.* **2022**; 226(1):167–176.
- 476 36. Hysmith ND, Kaplan EL, Cleary PP, Johnson DR, Penfound TA, Dale JB. Prospective
477 Longitudinal Analysis of Immune Responses in Pediatric Subjects After Pharyngeal
478 Acquisition of Group A Streptococci. *J Pediatr Infect Dis Soc.* **2017**; 6(2):187–196.
- 479 37. Walkinshaw DR, Wright MEE, Mullin AE, Excler J-L, Kim JH, Steer AC. The Streptococcus
480 pyogenes vaccine landscape. *Npj Vaccines.* Nature Publishing Group; **2023**; 8(1):1–6.
- 481
- 482
- 483
- 484
- 485
- 486
- 487
- 488
- 489
- 490
- 491
- 492

493 **Footnotes**

494

495 *Funding:*

496 This work was supported by a HEFCE/ODA grant from The University of Sheffield (155123)
497 and a Wellcome Trust Intermediate Clinical Fellowship award to TIdS [110058/ Z/15/Z]. The
498 funders had no role in study design, data collection and analysis, decision to publish, or
499 preparation of the manuscript.

500

501 *Conflicts of interest:*

502 We declare there are no conflicts of interest among the authors.

503

504 *Previous presentation:*

505 Preliminary data from this study were presented at the British Infection Society Spring
506 Meeting 2023, Manchester, United Kingdom. Data will also be presented at the European
507 Congress of Clinical Microbiology and Infectious Diseases 2024, Barcelona, Spain.

508

509 *Corresponding author:*

510 Jennifer Hall, Division of Clinical Medicine, School of Medicine and Population Health,
511 University of Sheffield, Sheffield, United Kingdom, Jennifer.Hall@sheffield.ac.uk